Title

Turning Logic Transistors into Secure, Multi-Time Programmable, Embedded Non-Volatile Memory Elements for 14 nm FINFET Technologies and Beyond

Author

Khan, Faraz

Publication Date

2019-06-01

Peer reviewed

Turning Logic Transistors into Secure, Multi-Time Programmable, Embedded Non-Volatile Memory Elements for 14 nm FINFET Technologies and Beyond

F. Khan[1], D. Moy[1], D. Anand[2], E. H.-Schroeder[2], R. Katz[1], L. Jiang[1], E. Banghart[1], N. Robson[1], T. Kirihata[1]
GLOBALFOUNDRIES ([1]NY USA, [2]VT USA), faraz.khan@globalfoundries.com

Abstract

Described is a secure, multi-time programmable memory (MTPM) solution for the 14 nm FINFET node and beyond, which turns as-fabricated standard logic transistors into embedded non-volatile memory (eNVM) elements, without the need for any process adders or additional masks. These logic transistors, when employed as eNVM elements, are dubbed "Charge Trap Transistors" (CTTs). Outlined are the technological breakthroughs required for employing logic transistors as an MTPM. An erase technique, called "Self-heating Temperature Assisted eRase" (STAR), is introduced which enables 100% erase efficiency, as compared to < 50% erase efficiency using conventional methods, in turn enabling MTPM functionality in CTTs. For the first time, hardware results demonstrate an endurance of > 10^4 program/erase cycles. Data retention lifetime of > 10 years at 125 °C and scalability to 7 nm have been confirmed.

Keywords: MTPM, CTTs, eNVM, self-heating.

Introduction

The need for on-chip non-volatile memory in VLSI technologies is ever-increasing. eFUSE and anti-fuse [1] technologies are one-time programmable, require high voltages which are logic incompatible, and programmed data can be reverse engineered using standard failure analysis techniques. Other memory technologies like MONOS [2] and MRAM [3], while multi-time programmable, require additional complex processes and masks. The CTT embedded non-volatile memory (eNVM) solution offers, without requiring any additional processes or masks, logic compatible operation voltages (~2V), multi-time programmability, data security (programmed data cannot be decoded using any currently known failure analysis techniques), and scalability.

CTTs (Fig. 1) are as-fabricated, standard high-k metal gate (HKMG) logic transistors [4], whose threshold voltages (V_T) may be modified by application of appropriate logic compatible voltages, where device self-heating enhanced charge trapping in the high-k gate dielectric ensures high data retention [5]. CTTs are programmed using short gate bias (V_G) pulses of 1.8-2.0V with a drain bias (V_D) of 1.4-1.6V, while the source bias (V_S) and the substrate bias (V_X) are at 0V. A 1.5Mb CTT one-time programmable memory (OTPM) product, capable of operating at military grade temperatures, has already been deployed [6]. However, poor erase efficiency - and consequent low program/erase (P/E) cycling endurance - has restricted the use of CTTs for multi-time programmable memory (MTPM) applications thus far. We introduce a technique that drastically improves the erase efficiency, and in turn, the cycling endurance of the CTT MTPM. For the first time, hardware results demonstrate an endurance of > 10^4 P/E cycles, a 1000× improvement, which is adequate for most embedded MTPM applications such as hardware security, encryption, firmware, chip ID, configuration, and repair.

"Self-heating Temperature Assisted eRase" (STAR)

Conventional erase operations (Fig. 2 (a)), typically performed using a negative gate bias (V_G) of magnitude > |2.5V|, while the source, drain and substrate are grounded, to electrostatically emit trapped charge, result in an inefficient erase (Fig. 3 (a)). Higher voltages cannot be used due to gate oxide breakdown concerns. The incomplete erase after each cycle causes the memory window to dynamically drift and become narrower, resulting in a shrinking read margin (Fig. 4 (a)). This severely limits the endurance (< 15 P/E cycles) and makes it challenging for implementation of CTTs as an MTPM technology, as circuits to dynamically change the reference current are difficult to implement. This problem is effectively addressed by using the STAR technique. Charge de-trapping in high-k dielectrics is strongly accelerated by temperature, usually defined by the Arrhenius model. The STAR technique (Fig. 2 (b)), utilizes the source-substrate-drain structure of the device as a parasitic NPN bipolar junction transistor (BJT) to pass a short current pulse through the body of the device during the erase operation. The device is biased such that, the parasitic BJT is in the active mode while there is a negative gate-to-substrate bias (V_{GX}) at the same time, without the need for any negative voltages. The local device self-heating caused by the BJT current, in combination with the negative V_{GX}, significantly enhances the charge de-trapping process: up to 100% erase efficiency (Fig. 3(b)) is achieved using lower voltages and shorter time as compared to the conventional erase method (100% erase within 1ms using STAR vs. < 50% even after 1s of conventional erase), in turn yielding a flat memory window with no narrowing for 10^4 P/E cycles (Fig. 4 (b)). 3D finite element thermal simulations of the respective bitcell temperatures during the erase operations performed using the two methods are also shown in Figs. 3 (a) and 3 (b). Simulation results estimate that steady state T is achieved within ~40-50 ns (Fig. 5). The measured I-V characteristics of the parasitic BJT are shown in Fig. 6.

The five-transistor 1.86 μm^2 STAR enabled MTPM bitcell design is shown in Fig. 7. In the memory array, each bitcell receives nine wires used to control or supply voltages generated on-chip from a 2.5V power supply during the modes of operation, shown in Fig. 7. The array is partitioned such that each wordline has a dedicated source line domain, which isolates the erase disturb (charge loss) to the bitcells on a common wordline. The bitcells that are exposed to the ~2V V_X and V_D (i.e. bitlines / columns on the same wordline) are sequentially erased due to this charge loss condition. However, cells on adjacent wordlines maintain a grounded V_X and V_D thereby avoiding erase disturb. The erase disturb isolation is done using four pFETs (Fig. 7) passing the voltages only to the row of memory cells that need to be erased. The bitcell pFETs are area efficient thin oxide devices, requiring stacking to keep transistors in safe operating regions through all modes of operation.

Data Retention

High-temperature charge retention bakes, performed on 14 nm FINFET CTTs (cycled using V_G=1.95V, V_D=1.55V for programming and erased using STAR), show a projected 10 year charge loss of < 25% at 125 °C (Fig. 8). The charge de-trapping activation energy (E_a), extracted using the conventional Arrhenius model, is ~1.85 eV. This is comparable to the reported E_a for one-time programmable 14 nm FINFET CTTs [6]. Differential sense current (ΔI_{SENSE}) distributions for a 9kb CTT array baked at 125 °C, for up to seven days (168 hours), are shown in Fig. 9.

Conclusions and Outlook

Demonstrated is the feasibility of an MTPM with > 10^4 P/E cycling endurance using CTTs in 14 nm FINFET technology as a secure eNVM solution for HKMG technologies that is logic voltage compatible and exhibits > 10 year data retention at 125 °C, without any added processes or masks. The predicted 14 nm CTT MTPM IP is 256kb/mm^2. Scalability of CTTs to 7 nm FINFET technologies has been confirmed: 100% erase efficiency and P/E cycling of 7 nm FINFET CTTs using STAR are shown in Fig. 10. Efforts towards implementation of CTT eNVM in FDSOI technology platforms are also ongoing.

This work is partly funded by the DMEA. We thank Professor S. S. Iyer of UCLA for his help and support.

References

[1] S.-Y. Chou, *ISSCC*, Feb. 2017, pp. 200-202. [2] S. Tsuda, *Proc. IEEE IEDM*, Dec. 2017, pp. 469-472. [3] K. Lee, *IEEE VLSI Technology Dig. Tech. Papers*, 2018, pp. 183-184. [4] J. Singh, *Symp. VLSI Technology*, 2017, pp. T140-T141. [5] F. Khan, *IEEE EDL*, Jan. 2016, pp. 88-91. [6] E. Hunt-Schroeder, *IEEE VLSI Circuits Dig. Tech. Papers*, 2018, pp. 87-88.

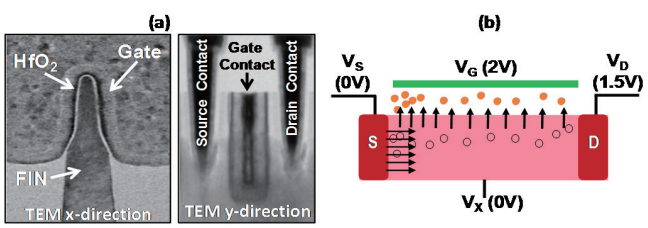


Fig. 1. 14nm FINFET CTT (a) TEM cross-sections in x- and y-directions, (b) schematic of programming operation.

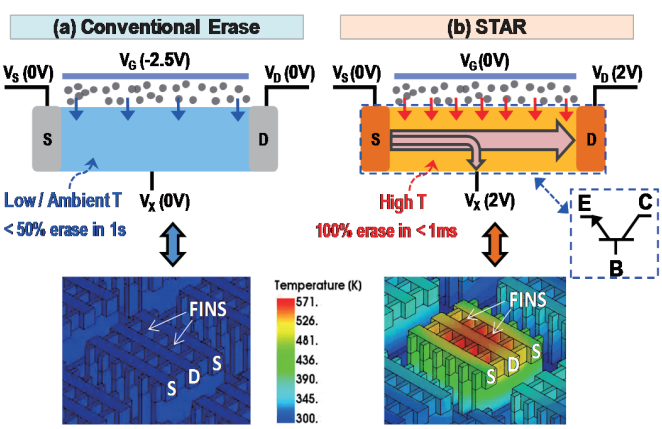


Fig. 2. Schematic showing (a) conventional erase and (b) "Self-heating Temperature Assisted eRase" (STAR). Corresponding thermal profiles of the bitcells during the erase operations are also shown for comparison.

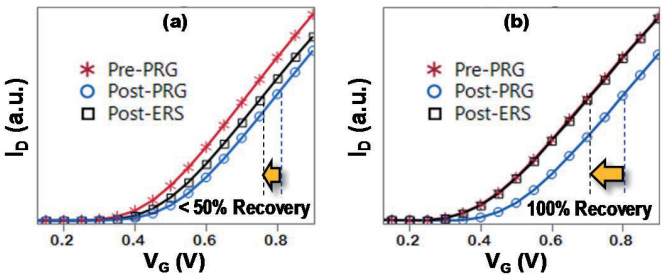


Fig. 3. Measured Pre-Program, Post-Program, and Post-Erase I_D-V_G data with (a) conventional erase and (b) STAR.

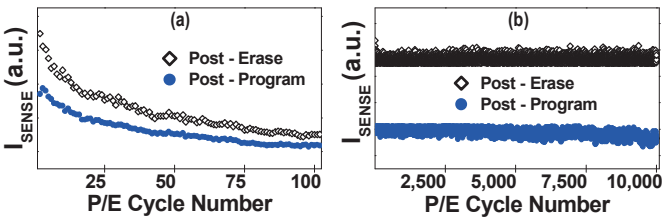


Fig. 4. P/E cycling of 14 nm FINFET CTTs using (a) conventional erase and (b) STAR.

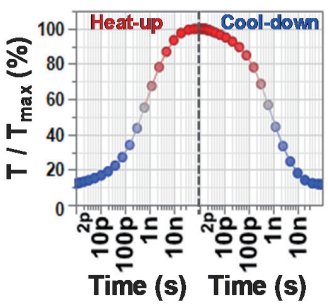


Fig. 5. Transient bitcell temperature vs. time for self-heating (during erase) and subsequent cool-down.

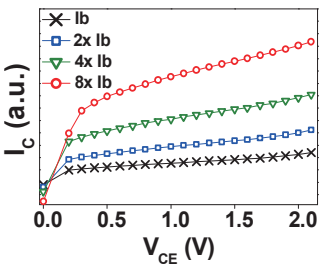
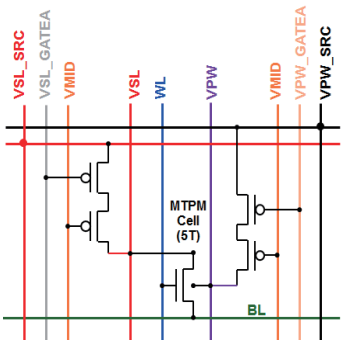


Fig. 6. Measured I-V characteristics of the parasitic BJT (described in Fig. 2 (b)).



Signal Voltages for Each Mode	Standby	Write Mask	Write Select	Erase Mask	Erase Select	Read
Voltages						
VSL_SRC	0	1.6	1.6	2	2	0
VSL	0	0	1.6	0	2	0
VPW_SRC	0	0	0	2	2	0
VPW	0	0	0	0	2	0
VMID	0.67	0.67	0.67	0.84	0.84	0.67
Gate controls						
VSL_GATEA	VDD	1.6	0.67	2	0.84	VDD
WL	0	0	1.9	0	0	~0.45
VPW_GATEA	VDD	VDD	VDD	2	0.84	VDD

Fig. 7. STAR enabled CTT bitcell design and typical operation conditions in 14 nm FINFET technology.

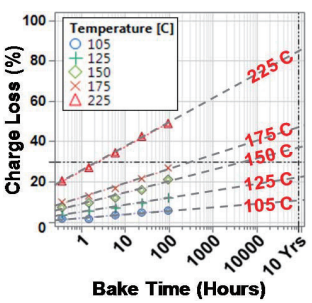


Fig. 8. High-temperature data retention bake tests.

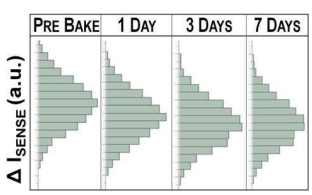


Fig. 9. ΔI_{SENSE} distributions in a 9kb CTT array baked at 125 °C, for up to 7 days.

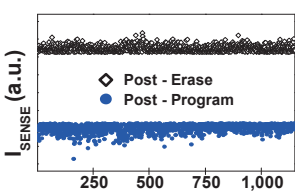


Fig. 10. P/E cycling of 7 nm FINFET CTTs using STAR.